

DOCUMENT RESUME

ED 177 642

CS 502 659

AUTHOR Trank, Douglas M.
TITLE The Effect of Student Feedback on the Improvement of Instruction.
PUB DATE [78]
NOTE 14p.; Research prepared at the University of Iowa
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Composition (Literary); *Educational Research; Effective Teaching; Evaluation Methods; *Feedback; Higher Education; *Student Evaluation of Teacher Performance; *Teacher Improvement

ABSTRACT

Fifty-four college rhetoric classes and their instructors were involved in an investigation of the extent to which student feedback gathered during the process of a course could improve instruction--or students' perception of that instruction--in a semester. At midterm, the classes were divided randomly into three groups. The first group of 18 were not administered a midterm evaluation by their instructors. During the eighth week of the semester, the Student Perceptions of Teaching (SPOT) form was administered to the 36 classes in the second two groups. Approximately a week after the instructors had administered these forms to their classes, each received a computer summary of individual results. The second group of instructors received only this printout, while the third group received the printout and also met with a faculty advisor to discuss the results. They also discussed the results with their respective classes. During the final week of classes, the same form was administered to all classes in the three groups. The results showed no significant differences between the two feedback groups nor between the two feedback groups and the control group. (FL)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

THE EFFECT OF STUDENT FEEDBACK ON THE IMPROVEMENT
OF INSTRUCTION

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Douglas M. Trank

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Douglas M. Trank
Associate Professor of Rhetoric and
Head of the Division of Speech Education
The University of Iowa
Iowa City, Iowa 52242

ED177642

The use of student evaluation of faculty is commonplace in post-secondary speech communication programs across the country. Although research on rating scales for such evaluation has been a topic of concern in educational journals for over fifty years, communication journals have generally avoided such discussions until the early part of this decade.¹ Wolff reported in 1974 that 90.5 per cent of the speech departments surveyed conducted such evaluations and all but one of the remaining respondents had plans to use them in the near future.² While we, as a discipline, have not been actively involved in the research surrounding student evaluation, it is obvious that most speech communication instructors do attempt to elicit some sort of systematic feedback regarding instruction from their students.

C 5502659

There has been and, in all probability, will continue to be a considerable amount of controversy surrounding the use of student ratings to assess instructor effectiveness in the classroom. Opponents have argued that the students' ratings are biased by the grades they receive, their age, sex, and field of concentration, the difficulty of the course, whether the course is required, and the amount of work required by the instructor. Such fears, according to a survey of research done by Larson, "appear to be unfounded."³ Such research indicates that student ratings of instructors are generally "reliable..., stable..., and valid."⁴ Cronen and Prince, however, did find differences in

undergraduate students' perspectives at different stages in their academic progress and urged educators to use student evaluation instruments with caution.⁵

Cronen later extended his research to graduate education where he found specific discriminations graduate students seemed to make in evaluating instructors.⁶ In a more recent article in this journal, Haslett demonstrated that "knowledgeability, or expertise of the student in a given area, did significantly influence that student's ratings of courses and instructors."⁷

Although the controversy certainly has not been resolved, it is not the purpose of this article to advance additional arguments on the topic of student rating validity and reliability. The focus, rather grows out of a statement offered two decades ago by Remmers:

No teacher has any choice as to whether he wishes to be judged by his students. The only choice he has is whether he wishes to know how he is judged and thus possibly capitalize on this feedback.⁸

What educators and administrators do with this information once it is collected is a potentially more valuable issue to pursue. In any situation where students are asked to comment on their instructors, both students and faculty need to be informed about the eventual application of such evaluations. Although we generally claim that the resulting data will be used by the instructors to improve their teaching, such evaluations are typically conducted at the end of the term when the instructors do not have an opportunity to alter either their behavior or the direction of their classes. The students are assessing something that has happened in the past, a product. The next time the instructor teaches that course, it will be a different class. The students, the context, the environment, and the communication situation will have changed. If we truly do want to utilize student opinion for the improvement of instruction, that input must come while the course is in process. Such "process" evaluations, as opposed to "product" evaluations, would allow for immediate adjusting by the instructors and the students.⁹

The question we have failed to answer fully is whether student feedback given during a course will have an effect on instructor behavior. More specifically, can student feedback gathered during the "process" of a course provide information useful for the improvement of instruction?

In a study involving 36 graduate student discussion leaders in three courses, Miller concluded that instructors given feedback at mid-term showed no difference in end-of-term ratings from instructors who were not given feedback.¹⁰ Centra reached a similar conclusion in a study which compared the effects of feedback over no feedback for instructors in all departments across five colleges.¹¹ Both Miller and Centra provided the instructors with printed summary reports with item means and figures and/or tables for making normative comparisons with their colleagues.

McKeachie and Lin conducted similar research with a 32-item rating scale administered by 37 graduate teaching assistants and three faculty members teaching in three psychology courses. One group in their study received the customary printed summary after a mid-term evaluation while a second group was provided with personal, verbal feedback and discussion. A third control group did not conduct mid-term evaluations. The personal feedback group was rated more effective by their students at the end of the semester than either of the other two groups.¹² The data in each of these studies were gathered from instructors teaching in a variety of courses. This confounds the issue not only because of the findings presented earlier, but because "it may be inappropriate to look for changes among a group of courses that may have different objectives and/or may employ different teaching techniques."¹³ Different courses have different objectives, goals, and methods of evaluation. Additional controls are needed to accurately assess the effect of such student feedback regarding instructor effectiveness. The following study was conducted to measure the effect of feedback to instructors who are all teaching the same course with the same objectives and using similar teaching techniques and student assignments.

METHOD

Fifty-four classes in the Rhetoric Program at the University of Iowa participated in the study. Fifty-one of the instructors were teaching assistants and three were full-time faculty. The program does not have a common syllabus and the teaching assistants are responsible for their own classes throughout the semester. The goals and objectives for the program, however, remain constant across sections and all members of the teaching staff attend regular teaching-learning sessions during their first semester as an instructor in the program. Thus, all instructors were teaching the same course (Freshman Rhetoric), using textbooks selected from a textbook list provided by the program, using similar speaking, reading, and writing assignments, and working toward the same objectives and goals.

At mid-term the classes were divided randomly into three groups. The first group (N=18) did not administer a mid-term evaluation to their classes. During the eighth week of the semester a 28-item Student Perceptions of Teaching (SPOT) form was administered to the 36 classes in the second two groups (N=19) (N=17). Twenty of the items had been chosen from a pool of over eighty items by a committee of full-time faculty and teaching assistants as being most appropriate for the subject matter and type of instruction found in the Rhetoric Program. The remaining items were written by this committee to obtain information specific to the teaching of speaking, writing, and critical reading not adequately covered in the existing item pool. All items call for the typical response ranging from Strongly Disagree (1) to Strongly Agree (6), and are stated positively so that a high rating is the most desirable outcome.

About a week after the instructors administered these forms to their students, they each received a computer summary of individual results providing the mean rating and three sets of percentiles for each item as well as the individual student response forms. The percentiles compared each instructor's class mean rating to three standards: 1) baseline data chosen to be representative of

courses at the University of Iowa, 2) accumulative data based on all instructors at the University who had used the item, and 3) accumulative data for the Rhetoric Program which had been gathered on these items for the four previous semesters (over 10,000 students). The second group of instructors (N=19) received only this printed feedback from our computer center. The third group of instructors (N=17) also received this printed feedback. In addition, operating on the "clear evidence that student opinions are most helpful when teachers discuss these responses with a colleague,"¹⁴ this group also met with a faculty advisor to discuss these results. After this discussion the instructors then discussed the results with their respective classes. The object of these discussions was to help the instructors respond directly to the evaluations of their students. The discussions with their students was intended to clarify the student ratings, find out how the students thought the instructors could improve their teaching and their courses, and, in turn, improve the student ratings of the instructors by the end of the semester. During the last week of classes of that semester, the same 28-item rating form was administered to all classes in the three groups.

RESULTS AND DISCUSSION¹⁵

Because each item on the SPOT form is intended to provide feedback on a specific characteristic of teaching, the analyses dealt with the items individually. The data were the average class ratings of each of the 28 items for each class participating in the study. Table 1 displays the average class ratings for each item in the three groups with both the mid-term and end-of-semester means for the two feedback conditions.

Insert Table 1 about here

To test whether or not any change occurred from mid-term to the end of the term, a two factor analysis-of-variance with repeated measures on one factor¹⁶ was computed for each of the 28 items in the two feedback conditions. Table 2 displays the results of the analyses for all the items.

Insert Table 2 about here

Since we have run 28 separate analyses, it is appropriate that we should consider an alpha level of significance less than $p < .05$, which we might choose for a single analysis. Accordingly, if a significance level of $p < .002$ is applied to these data, no item showed any significant difference from mid-term to end-of-term for the two feedback groups.

The next analysis considered whether there were any differences in end-of-term ratings among all three groups. The assumption was that the two feedback groups would show more agreement among the students than the groups which had no mid-term feedback. Table 3 displays the results of a two factor analysis-of-variance in which the groups are the between-subject effect and the items

Insert Table 3 about here

are the within-subject effect. The mean ratings for the eight items appearing on the back of the form were not included in this analysis since in two classes of the third group the students did not respond to some of the items. As Table 3 indicates there was no significant difference among the three groups at the end of the term.

Though the groups did not differ at the end of the semester at any appropriate statistical level of significance, looking again at Table 1 we did note

an interesting pattern in the end-of-semester item means. In general, the no-feedback group had lower ratings on more of the items than the end-of-term ratings for either of the other two groups. Twenty of the 28 items in the verbal feedback group had higher ratings than did the corresponding items in the no-feedback group. Statistically the chances of that occurring if feedback had no effect (using the normal approximation to the binomial distribution) are less than 5%. The printed feedback group had higher ratings than the no feedback group on 23 of the items ($p < .001$). These findings indicate that the perceptions of the students in the two feedback conditions are more positive at the end of the semester than the students whose instructors received no feedback at mid-term. Though the mid-term ratings of the two feedback groups also tended to be higher than the end-of-term ratings for the no feedback group, there is no reason to believe that at the outset the groups were different since they were chosen by random selection. This does suggest that the end-of-term ratings in the no-feedback group may be lower than the ratings that these instructors might have received at mid-term.

Neither type of feedback produced statistically significant differences in the end of the semester ratings for the three groups. This is a disappointing conclusion for those who believe they can use systematically gathered student feedback to improve their instruction, or at least their student's perception of their instruction. It may be that instructors are incapable of interpreting student responses in order to adjust their behavior and teaching or that they are incapable of making those adjustments even when they can interpret the students' opinions. It may be that it is inappropriate to use the same instrument to gather information that will be used for process and product appraisal applications. This research does indicate that we need to be extremely cautious in proclaiming that we gather student feedback regarding instructors and courses for the improvement of instruction.

FOOTNOTES

1. Florence I. Wolff, "Student Evaluation of College and University Speech Communication Courses and Faculty: A Survey," The Speech Teacher, 23 (September 1974), 252.
2. Wolff, 252-253.
3. Richard Larson, The Evaluation of Teaching College English (New York: MLA/ERIC, 1970), 21.
4. Larson, 21.
5. Vernon E. Cronen and William K. Price, "Class Year, Dimensions of Student Judgment, and the Use of Course Evaluation Instruments," The Speech Teacher 23 (January 1974), 34-39.
6. Vernon Cronen, "Course Evaluation in the Graduate Program," Communication Education, 25 (March 1976), 138-147.
7. Betty J. Haslett, "The Influence of Student Knowledgeability on Student Ratings of Instruction," Communication Education, 26 (January 1977), 44-51.
8. H. H. Remmers, "Students Perceptions of Teacher Effectiveness," in W. J. McKeachie, Ed., The Appraisal of Teaching in Large Universities (Ann Arbor: University of Michigan Press, 1958), 20.
9. For an excellent discussion of the distinction between process and product evaluation, see Gary D. Borich, The Appraisal of Teaching: Concepts and Process (Reading, Mass: Addison-Wesley Publishing Company, 1977), 36-42. I am using process evaluation to refer to what Borich calls formative evaluation-- "the identification of absent or weak skills and the recommendation of instructional experiences designed to help the teacher attain or improve these skills." (36) Product or summative evaluation, on the other hand, is concerned with providing data which can be "Used to form cumulative judgments about the teacher, incorporating a broad range of teacher performance variables." (37) Although we generally claim to use student evaluations to improve instruction, such data is most often relegated to summative appraisal applications influencing decisions about merit salary increases, promotion, tenure, and reemployment. Whether the same data can effectively be applied for both formative and summative objectives is a critical question that is yet to be answered.
10. M. T. Miller, "Instructor Attitudes Toward, and their Use of, Student Ratings of Teachers," Journal of Educational Psychology, 62 (December 1973), 395-401.
11. J. A. Centra, "The Effectiveness of Student Feedback in Modifying College Instruction," Journal of Educational Psychology, 65 (June 1971), 235-239.

12. W. McKeachie and Y. G. Lin, "Use of Student Ratings in Evaluation of College Teaching," Final Report, (National Institute of Education), University of Michigan, 1975. (Mimeo).
13. Rina R. Werts, Student Perceptions of Teaching (SPOT): The Use of Feedback From Student Ratings For Improving College Teaching, University of Iowa, 1978. (Mimeo).
14. "Using Student Reactions," Memo To The Faculty. Ann Arbor: University of Michigan, 58 (October 1976), 5.
15. The author would like to acknowledge the contributions made in this section by Rina Weerts of the Evaluation and Examination Service at the University of Iowa.
16. B. J. Winer, Statistical Principles in Experimental Design (New York: McGraw-Hill, 1971).

Table 1

Average Class Ratings for Three Groups

Item # ^a	Group I No Feedback N=18	Group II Printed Feedback N=17		Group III Verbal Feedback N=19	
		Mid	End	Mid	End
103	4.36	4.48	4.67	4.38	4.44
115	4.67	4.65	4.77	4.79	4.78
123	4.71	4.77	4.90	4.82	4.84
203	4.62	4.74	4.83	4.72	4.79
207	4.90	4.99	4.92	5.00	4.97
209	5.23	5.46	5.38	5.37	5.37
233	4.80	4.39	4.86	4.74	4.74
234	5.05	5.13	4.97	5.06	5.01
239	5.36	5.41	5.33	5.42	5.27
241	4.80	4.84	4.87	4.89	4.90
317	4.64	4.68	4.83	4.69	4.77
321	4.74	4.88	4.79	4.81	4.79
326	4.72	4.86	4.86	4.78	4.70
333	4.58	4.65	4.78	4.68	4.73
405	4.50	4.46	4.47	4.39	4.50
407	4.68	4.68	4.89	4.69	4.76
409	4.66	4.58	4.85	4.60	4.71
413	4.67	4.71	4.80	4.74	4.83
414	4.57	4.51	4.56	4.58	4.55
415	4.62	4.55	4.75	4.74	4.67
Opt 1	5.12	4.98	4.97	5.29 ^b	5.09 ^b
Opt 2	4.62	4.70	4.85	4.78 ^b	4.81 ^b
Opt 3	4.88	4.93	5.00	4.96 ^b	4.97 ^b
Opt 4	4.87	4.93	5.13	4.82 ^b	4.93 ^b
Opt 5	3.71	4.15	3.99	4.15 ^b	3.89 ^c
Opt 6	5.08	5.06	5.19	4.94 ^b	4.79 ^c
Opt 7	4.86	5.07	5.15	4.95 ^b	4.99 ^c
Opt 8	3.86	4.09	4.18	4.15 ^b	4.04 ^c

^aFor item texts, see Appendix A

^bN=18

^cN=17

Table 2

ANOVA of Average Class Means of Mid-Term and End-of-Term
Ratings for Two Feedback Groups^a

Item # ^b	Mean Square Mid-End (df)	Mean Square Mid-End x Group (df)	Mean Square Error (df)	F Interaction
103	0.2580 (1)	0.0717 (1)	0.0653 (34)	1.0989
115	0.0435 (1)	0.0781 (1)	0.0515 (34)	1.5165
123	0.0990 (1)	0.0590 (1)	0.0942 (34)	0.6266
203	0.1160 (1)	0.0005 (1)	0.0739 (34)	0.0068
207	0.0501 (1)	0.0041 (1)	0.0979 (34)	0.0423
209	0.0276 (1)	0.0317 (1)	0.0703 (34)	0.4507
233	0.0031 (1)	0.0034 (1)	0.0926 (34)	0.0370
234	0.1995 (1)	0.0591 (1)	0.0542 (34)	1.0909
239	0.2369 (1)	0.0169 (1)	0.1209 (34)	0.1396
241	0.0052 (1)	0.0030 (1)	0.0671 (34)	0.0447
317	0.2245 (1)	0.0237 (1)	0.1269 (34)	0.1865
321	0.0512 (1)	0.0224 (1)	0.1121 (34)	0.1994
326	0.0334 (1)	0.0315 (1)	0.0693 (34)	0.4546
333	0.1318 (1)	0.0270 (1)	0.0554 (34)	0.4865
405	0.0793 (1)	0.0440 (1)	0.1112 (34)	0.3959
407	0.3227 (1)	0.0977 (1)	0.0863 (34)	1.1317
409	0.6142 (1)	0.1220 (1)	0.0812 (34)	1.5021
413	0.1440 (1)	0.0004 (1)	0.0798 (34)	0.0056
414	0.0004 (1)	0.0281 (1)	0.1141 (34)	0.2459
415	0.0517 (1)	0.3153 (1)	0.0606 (34)	5.2046*
Opt 1	0.0272 (1)	0.0085 (1)	0.0909 (32)	0.0935
Opt 2	0.1650 (1)	0.0420 (1)	0.1314 (32)	0.3197
Opt 3	0.0577 (1)	0.0068 (1)	0.1039 (32)	0.0654
Opt 4	0.5948 (1)	0.0043 (1)	0.1299 (32)	0.0330
Opt 5	0.6005 (1)	0.0150 (1)	0.1552 (32)	0.0967
Opt 6	0.0032 (1)	0.2436 (1)	0.1254 (32)	1.9423
Opt 7	0.1271 (1)	0.0001 (1)	0.0866 (32)	0.0011
Opt 8	0.0000 (1)	0.1487 (1)	0.0852 (32)	1.7447

*p < .05

^aVerbal Feedback Group N = 19 (N = 17 for optional items)

^bFor item texts, see Appendix A

Table 3
ANOVA on End-of-Term Ratings for
3 Groups

Source of Variation	SS	df	MS	F
Between Classes	219.01	53	4.13	
Groups	8.79	2	4.40	1.07
Err (Between)	210.22	51	4.12	
Within Classes	156.18	1026	0.15	
Items	53.13	19	2.80	27.17
Items x Groups	3.35	38	0.09	0.86
Err (Within)	99.71	969	0.10	
Total	375.19	1079	0.34	

Appendix A

115. Course objectives helped me understand main emphases.
 103. Subject matter was intellectually stimulating to me.
 123. Instructor emphasized ways of solving problems rather than solutions.
 209. Instructor seemed interested in teaching this course.
 239. Instructor encouraged students to see him/her if they were having difficulty.
 207. Instructor seemed to be concerned with whether I learned the material.
 233. Instructor invited criticisms of his/her own ideas.
 234. I was encouraged to participate in class discussion.
 241. Instructor discussed points of view other than his/her own ideas.
 203. My work was evaluated in ways that were meaningful to me.
 333. Instructor had me apply concepts to demonstrate understanding.
 321. Assigned readings were pertinent to topics presented in class.
 326. Discussion was helpful to my learning.
 317. Assignments contributed to my learning.
 405. I was stimulated to discuss new ideas in or out of class.
 414. Instructor helped me feel confident in expressing new ideas.
 409. My skill in critical thinking was increased.
 415. I learned to see relationships among important topics and ideas.
 413. I learned to understand my strengths and weaknesses in the area.
 407. My knowledge and skills were increased.
-
1. Group interaction was a part of the class.
 2. The teacher's explanations of concepts were valuable.
 3. The instructor effectively integrated reading, writing, and speaking.
 4. There was an appropriate amount of work for credit earned.
 5. The textbooks were useful to the course.
 6. The instructor returned evaluations of my work in speaking, reading, and writing with reasonable promptness.
 7. Generally, I would rate my instructor above average.
 8. Generally, I would rate this course above average.